

## **AMENDMENTS TO THE SPECIFICATION**

**With the section starting on page 5, line 2, please amend the specification as follows:**

Figure 1A shows amounts of cDNA after two rounds of T7 amplification.

Figure 1B shows percent transcripts detected in normal and tumor tissues.

Figure 2A shows 39 genes whose expression changed in all five samples taken from patients participating in a pilot cancer study.

Figure 2B shows a representative sample of the differentially expressed genes grouped into biological pathways known to be relevant in carcinogenesis

Figure 2C shows candidate genes that are up regulated.

Figure 2D shows candidate genes that are down regulated.

Figure 3 shows a comparison of percent increases for three upregulated genes measured by ~~GeneChip®~~ GENECHIP® and Real Time Quantitative PCR data.

Figure 4 shows a comparison of differential collagenase gene expression measured by ~~GeneChip®~~ GENECHIP® microarray and RT-QPCR.

Figure 5 shows hierarchical clustering.

Figure 6 shows differentially expressed genes identified by three different methods.

Figure 7 shows differential gene expression using ~~GeneChip®~~ GENECHIP® analysis software which revealed that 404 genes are differentially expressed.

**With the paragraph starting on page 17, line 19, please amend the specification as follows:**

The patent also shows a system block diagram of a computer system used to execute the software of an embodiment of the invention. The computer system includes a monitor, a keyboard, and a mouse. The computer system further includes subsystems such as a central processor, a system memory, a fixed storage (e.g., a hard drive), a removable storage (e.g., CD-ROM), a display adapter, a sound card, speakers, and a network interface. Other computer systems suitable for use with the invention may include additional or fewer subsystems. For example, another computer system may include more than one processor or a cache memory. Computer systems suitable for use with the invention may also be embedded in a measurement instrument. The embedded systems may control the operation of, for example, a ~~GeneChip®~~ GENECHIP® Probe array scanner (also called a GENEARRAY® ~~GeneArray™~~ scanner sold by Agilent corporation, Palo Alto Ca.) as well as executing computer codes of the invention.

**With the paragraph starting on page 18, line 10, please amend the specification as follows:**

Computer software to analyze data generated by microarrays is commercially available from Affymetrix Inc. (Santa Clara) as well as other companies. Affymetrix Inc. distributes ~~GeneChip®~~ GENECHIP®, now known as MicroArray suite, LIMS, Microdb, Jaguar, DMT, and other software. Other databases can be constructed using the standard database tools available from Microsoft (e.g., Excel and Access).

**With the paragraph starting on page 18, line 26, please amend the specification as follows:**

High-density oligonucleotide arrays are particularly useful for monitoring the gene expression pattern of a sample. In one approach, total mRNA isolated from the sample is converted to labeled cRNA and then hybridized to an array such as a ~~GeneChip®~~ GENECHIP® oligonucleotide array. Each sample is hybridized to a separate array. Relative transcript levels are calculated by reference to appropriate controls present on the array and in the sample.

**With the paragraph starting on page 19, line 19, please amend the specification as follows:**

Normal and tumor cells from a solid tumor site from within the oral cavity were obtained using laser capture microdissection as described in provisional application 60/182,452 (attorney docket number 3294) which is hereby incorporated by reference in its entirety for all purposes. According to that method, biopsies were taken and snap frozen. The biopsies were sectioned at 5 microns and mounted on slides. They were then stained with hematoxylin and eosin. Laser capture microdissection was then used to procure malignant and normal keratinocytes. Laser capture microdissection, RNA isolation, IVT, 3 rounds of T7 RNA polymerase linear amplification, and probe biotinylation were carried out according to the methods of Alevizos et al., submitted, (2000) and Ohyama et al., Biotechniques 29, 530-6 (2000), each of which are hereby incorporated by reference in their entireties for all purposes. Basically, RNA was extracted and then cDNA synthesis was carried out using Superscript SUPERSCRIPT™ (Life Technologies). cRNA synthesis and labeling was carried out using Ampliscribe (Epicenter technology) and BioArray High Yield RNA Transcript Labeling System (Enzo).

**With the paragraph starting on page 20, line 3, please amend the specification as follows:**

The quality and quantity of isolated RNA was examined by reverse transcription polymerase chain reaction (RT-PCR) of five cellular maintenance gene transcripts of high to low abundance (glyceraldehyde-3-phosphate dehydrogenase, tubulin- $\alpha$ ,  $\beta$ -actin, ribosomal protein S9, and ubiquitin C) (Ohyama et al., 2000). The quantity of isolated RNA was also assessed with ~~RiboGreen~~ RIBOGREEN® RNA Quantitation Reagent and kit (Molecular Probes, Eugene, OR) using spectrofluorometry (Bio-Rad, Hercules, CA). Only those samples exhibiting PCR products for all five cellular maintenance genes were used for subsequent analysis. The biotinylated cRNA from the ten samples (five normal and five cancer) were further used to hybridize the Affymetrix Test-1 probe arrays to determine cRNA quality and integrity. The arrays contain probes representing a handful of maintenance genes and a number of controls (Ohyama et al., 2000). Analysis of the arrays confirmed the RT-PCR findings. cRNA linearly amplified from human oral cancer tissue produced no nonspecific or unusual hybridization patterns, and transcripts corresponding to the maintenance genes were detected. The 5' region of the RNA was degraded, but sufficient 3' transcript was intact to proceed with hybridization using the HuGeneFL probe arrays. In addition, probes synthesized on the arrays are biased to the last 600 bp in the 3' region of the transcripts. Yields of cDNA resulting from the LCM, RNA isolation, and after two rounds of T7 amplification are shown in Figure 1A. Linear amplification of total RNA began with ~100 ng of total RNA. As shown in Figure 1A, the amount of double stranded cDNA (ds-cDNA) after two rounds of T7 amplification is dependent on the quality of the LCM-generated RNA from the normal and tumor tissues.

**With the paragraph starting on page 21, line 3, please amend the specification as follows:**

The cRNA was fragmented as described by Wodicka et al. (1997) and then hybridized to Affymetrix probe arrays such as ~~GeneChip~~ GENECHIP® Test 1, Human U95A and HuGeneFL probe arrays. Hybridization was carried out for a time period of between about 12 to 16 hours. All array washing, staining and scanning were carried out as described in the Gene Expression Manual (Affymetrix, Inc. 1999 hereby incorporated by reference in its entirety for all purposes). The Affymetrix arrays include probe sets consisting of oligonucleotides 25 bases in length. Probes are complementary to the published sequences (~~GeneBank~~ GENBANK®) as previously described (Lockhart et al., 1996). The sensitivity and reproducibility of the ~~GeneChip~~ GENECHIP® probe arrays is such that RNAs present at a frequency of 1:100,000 are unambiguously detected, and detection is quantitative over more than three orders of magnitude

(Redfern et al., 2000; Warrington et al., 2000). In this set of experiments with oral cancer samples, the bacterial transcript (BioB), spiked before the hybridization at concentration of 1.5 pM. This concentration, which corresponds to three copies per cell, based on the assumption that there are 300,000 transcripts per cell with an average transcript length of 1 kb, was defined as present in nine out of ten experiments (Lockhart et al., 1996). Array controls, and performance with respect to specificity and sensitivity are the same as those previously described (Lockhart et al., 1996; Mahadevappa & Warrington, 1999; Wodicka et al., 1997). Information regarding the genes represented on the arrays used in this experiment can be found at [www.netaffx.com](http://www.netaffx.com).

**With the paragraph starting on page 21, line 26, please amend the specification as follows:**

The data obtained from the microarrays was analyzed using various methods and software commercially available and known to those skilled in the art. These methods and software include T-test, ~~GeneChip®~~ GENECHIP® software available from Affymetrix to perform a comparison analysis; GeneCluster SOM software to perform a cluster analysis, identify genes and develop characteristics of gene expression profiles; and ~~Matlab~~ MATLAB™ software to identify genes that are differentiating and to identify gene classes. Additional software that can be used to analyze chip data includes GenExplore and PCA.

**With the paragraphs starting on page 22, line 3, please amend the specification as follows:**

For GeneCluster analysis and the computation of self organizing maps (SOM), gene expression levels and geometry of nodes were input into the GeneCluster software. Before the computation of the SOM, two preprocessing steps took place. First, a filter was applied to exclude genes that did not change significantly across the pairs. Genes were eliminated if they did not show a relative change of  $x=2$  and an absolute change of  $y=35$ ,  $(x, y)=(2, 35)$ . Second, normalization of expression levels across experiments was carried out, thus emphasizing the expression pattern rather than the absolute expression values. Data was normalized using ~~GeneChip®~~ GENECHIP® software. A description of the normalization procedure can be found on pp. A5-14, ~~GeneChip®~~ GENECHIP® Expression Analysis Technical Manual, (Tamayo et al., 1999).

Differential gene expression using ~~GeneChip®~~ GENECHIP® analysis software revealed that 404 probe sets changed in the majority of the cases (3/5) (set forth in Figure 7). Among the 404, 211 were increased in tumor samples and 193 were decreased in tumor samples, compared to normal samples. As shown in Figure 2A, 39 probe sets used allowed the detection of changes in gene expression in all five cases. Sixteen genes showed increased expression in tumor

samples and 23 genes showed decreased expression in tumor samples, compared to normal samples. Figure 2B is a list of differentially expressed genes grouped into biological pathways known to be relevant in carcinogenesis. Figure 2C is a list of differentially expressed genes which are up regulated in tumor samples. Figure 2D is a list of differentially expressed genes which are down regulated in tumor samples.

**With the paragraphs starting on page 23, line 8, please amend the specification as follows:**

In order to validate the expression level data, three metastatic pathway genes whose expression are consistently altered in the five paired cases of oral cancer were selected. Real-time quantitative PCR (RT-QPCR) in conjunction with the ~~TaqMan~~ TAQMAN® specific probe system or SYBR® Green system were used to validate the expression levels of interstitial collagenase (a member of the MMP's involved in metastasis), urokinase plasminogen activator (UPA, associated with metastasis) and cathepsin L (a member of the serine proteases). The cDNA product of the reverse transcription reaction was used as the template for the RT-QPCR reaction. For the RT-QPCR reaction, the iCycler IQ™ Real Time PCR detection system (Bio-Rad, Hercules, CA) was used with ~~TaqMan~~ TAQMAN® specific probes and primers for Cathepsin, and SYBR® Green buffer and reagents (Perkin Elmer/Applied Biosystems Foster City, CA, USA) for Urokinase Plasminogen Activator and Collagenase I (Heid et al., 1996).

For designing the specific primers and probes, PE/ABD Primer Express software as well as ~~MaeVector~~ MACVECTOR® were used. Primer sequences used were:

Collagenase forward: 5'-ACACGGAACCCCAAGGACA-3' (SEQ ID NO:1)

Collagenase Reverse: 5'-GTTTTGTTGCCGGTGGTTTT-3' (SEQ ID NO:2)

UPA forward: 5'-GCACCATCAAACAAACCCCTTAC-3' (SEQ ID NO:3)

UPA reverse: 5'-CAGACAGAAAAACCCCTGCCTG-3' (SEQ ID NO:4)

Cathepsin L forward: 5'-CAGTGTGGTTCTTGTGGGGCT-3' (SEQ ID NO:5)

Cathepsin L reverse: 5'-CTTGAGGCCAGAGCAGTCTA-3' (SEQ ID NO:6)

**With the paragraph starting on page 23, line 30, please amend the specification as follows:**

Comparison of the microarray and RT-QPCR data as shown in Figure 3 revealed that they approximate each other. The slight observed discrepancy in the precise quantitation of the ~~GeneChip®~~ GENECHIP® and the RT-QPCR was due to the fact that a minute amount (ng) of LCM-generated total RNA was used for amplification followed by biotinylation and hybridization to the ~~GeneChip®~~ GENECHIP® microarrays. Using the same LCM-generated total RNA, the ~~GeneChip®~~ GENECHIP® data of three metastatic tumor genes was validated by real-

time quantitative PCR (RT-QPCR). These two independent approaches yielded data which indicated a similar trend (Figure 4). Both methods indicate that genes were upregulated from undetectable levels in the control to moderate abundance in the tumor cells. Similar results of ~~GeneChip®~~ GENECHIP® versus RT-QPCR correlation were previously used by Welsh et al. to validate candidates identified in an ovarian cancer study (Welsh et al., 2001). The RT-QPCR data confirmed the upregulation and downregulation of selected candidates. Therefore, while there is discrepancy in the precise quantitation of ~~GeneChip®~~ GENECHIP® and RT-QPCR data of each sample, the overall trend and correlation are similar. The array data produces information about relative abundance that is accurate to within 1.5 to 2 fold (Lockhart et al., 1996; Redfern et al., 2000) providing information that allows binning of the transcript levels by low, low-medium, medium, medium-high or high abundance (Warrington et al., 2000a; Warrington et al., 2000b; Lockhart et al., 1996; Redfern et al., 2000).

**With the paragraph starting on page 25, line 23, please amend the specification as follows:**

Using ~~Matlab~~ MATLAB™ software analysis, 117 transcripts were identified to be differentially expressed between normal and tumor cells. Hierarchical clustering is shown in Figure 5. The distinct clustering of the normal samples from the tumor samples suggests that LCM procured pure, homogenous samples.

Based on the outcome of three analytical methods (~~GeneChip®~~ GENECHIP®, SOM and ~~Matlab~~ MATLAB™), ~600 candidate oral cancer genes were identified. Of this comprehensive set, 27 of the differentially expressed genes were identified by all three methods (set forth in Figure 6). Of the 600 candidate genes, 41% were detected at low levels, 1-5 copies per cell.

**With the paragraphs starting on page 26, line 3, please amend the specification as follows:**

Shillitoe *et al.* and Leethanakul *et al.* have created expression libraries of human oral cancer cell lines and LCM-generated oral cancer tissues (Leethanakul et al., 2000a; Leethanakul et al., 2000b; Shillitoe et al., 2000). Their studies revealed 52 genes to be differentially expressed at >2-fold in at least three of the cancer tissue sets. Of these 52 genes, 26 were present on the Affymetrix ~~GeneChip®~~ GENECHIP®. Of these 26 overlapping genes, 18 were called absent (not detectable) in both normal and tumor samples (DP-2/U18422; TIMP-4/U76456; VEGF-C/U43142; FGF3/X14445; FGF5/M37825; FGF6/X63454; IGFBP5/M65062; EGF cripto protein CR1 and 2/M96956; APC/M74088; ERK6/X79483; GDI dissociation protein/U82532; MAP kinase p38/L35253; MKK6/U39657; MEKK3/U78876; Frizzled/L37782; FZD3/U82169; Dishevelled homolog/U46461; Patched homolog/U43148;); one gene showed no difference

between normal and tumor tissues (cyclin H/U11791); one gene was upregulated in five out of five tumors (beta1-catenin/X87838); three genes were upregulated in four out of five tumors (thrombospondin2 precursor/L12350; inhibitor of apoptosis protein/U45878; Caspase 5 precursor/U28014); one gene was upregulated in three out of five tumors (MMP-10/X07820); and one gene was downregulated in four out of five tumors (RhoA/L25080). Finally, one gene was upregulated in two tumors, downregulated in two tumors, and called absent in the fifth oral cancer (TRAF2/U78798).

Of the 52 genes, two genes were detected present only through LCM/~~GeneChip®~~ GENECHIP® analysis. They are human SPARC/osteonectin (J03040) and 5T4 oncofetal antigen (Z29083), which are consistently altered in the same manner in all five oral cancers examined.

**With the paragraph starting on page 27, line 9, please amend the specification as follows:**

The different outcomes of the various studies are likely reflective of the experimental approaches and methods of analyses. First, by using LCM-generated RNA, contamination of heterogeneous cellular elements is avoided. Second, sample number and the type of microarray used in the respective studies may be relevant to the discrepancies. Third, the stage of the tumor, source and anatomical site of the oral cancers, and handling methods can further result in different gene expression levels. However, LCM-generated RNA, linearly amplified by T7 RNA polymerase and subsequently analyzed by high-density oligonucleotide ~~GeneChip®~~ GENECHIP® probe arrays impressively provided for the detection of 39 cellular genes consistently altered in five out of five different paired cases of human oral cancer making these genes useful as classifiers to predict the normal/malignant nature of oral epithelial tissues.